구매자의 탐색 패턴에 기반한 상품 검색 지원 에이전트의 설계 및 구현

홍영준[†] · 박경환[††]

요 약

본 논문에서는 인터넷 쇼핑몰에서 구매자가 효과적으로 상품을 검색할 수 있는 상품 검색 지원 에이전트를 설계하고 구현하였다. 기존 쇼핑몰에서의 전자 카탈로그는 주제별 또는 검색어에 의한 상품 정보를 제공하거나 상품구매 이력에 관한 구매 정보를 구매자에게 제시하였다. 본 논문에서는 보다 효과적인 구매 정보를 제공하기 위해 기존 구매자의 탐색 패턴을 분석하여 이를 근거로 하여 구매하려는 상품과 관련된 상품 정보를 제시해 주는 상품검색 지원 에이전트를 개발하였다. 따라서 구매자는 구매하려는 상품과 관련된 정보를 보다 효과적으로 얻을 수 있어 효율적인 쇼핑을 즐길 수 있을 것이다.

A Design and Implementation of Product Search Support Agent based on the Behavior of Customers

Young-Jun Hong[†] and Kyung-Hwan Park[††]

ABSTRACT

This paper introduces the design and implementation of product search support agent for an efficient Internet shopping. In existing shopping malls, electronic catalogs provide the product information by subject or keyword, or the buying information by purchasing history. This proposed product search support agent system is designed to analyze navigating patterns of customers, extract more effective buying information with regard to the current trend in customer navigating. In conclusion, this product search support agent can propose more suitable buying information to customers that he can be proposed to make purchases more efficiently.

1. 서 론

최근 인터넷의 폭발적인 확산에 따라 전자상거래가 시간적, 공간적 제약을 극복한 새로운 시장으로 부각되고 있다. 전자 상거래는 기업, 개인, 정부 등 경제 주체들이 정보 통신 기술을 활용하여 상품 및 서비스를 교환하는 거래 방식이다. 전자 상거래의 유형은 경제 주체간의 관계에 따라 설정할 수 있으며, 그 중 기업과 개인간의 관계가 인터넷 쇼핑몰의 근간을 이룬다고 할 수 있다[1].

전자상거래 시스템은 크게 고객 시스템, 상점 시스템, 지불 시스템 및 인증 시스템으로 나눌 수 있다[2]. 그 중 상점 시스템에는 고객에게 상품 정보를 제공하고 사용자와의 인터페이스 역할을 하는 쇼핑몰이 있다. 전자 쇼핑몰은 구매자에게 효과적인 상품 정보를 제공하기 위해 다양한 전자 카탈로그를 제공한다. 기존의 쇼핑몰에서는 상품 정보에 대한 주제별 검색 또는 검색어에 의한 상품 정보를 제공하고 있는 것이 일반적이다. 또한 보다 효과적인 상품정보를 제공하기 위해 특정 상품을 검색하는 구매자에게 해당 상품의 판매 이력을 제시함으로써 보다 효과적인 상품 정보를 제공하기도 한다. 그러나 구매자에게 보다

본 연구는 한국과학재단 지정 동아대학교 지능형통합항만관리연구센터의 지원에 의한 것입니다.

† 동아대학교 컴퓨터 공학과 석사 과정
†† 동아대학교 전기전자컴퓨터컴퓨터공학부 교수

효과적인 구매 정보를 제공하려면 구매자의 구매 의도를 파악하여 보다 효과적인 구매정보를 제시하는 방법의 추구가 필요하다[3].

본 논문에서는 인터넷 쇼핑몰에서 보다 효과적인 구매 정보를 제시하기 위해 기존 구매자의 탐색 패턴을 분석하여 이를 근거로 하여 구매자가 구매하려는 상품과 관련된 상품 정보를 제시해 주는 상품 검색 지원 에이전트를 설계하고 구현하였다. 따라서 본 검색 지원 에이전트를 사용하는 시스템은 구매자에게 보다 효과적인 구매 정보를 제공할 수 있어 구매자는 보다 효율적인 쇼핑을 할 수 있을 것으로 사료된다. 본 논문에서는 상품 검색 지원 에이전트의 구축 영역으로 인터넷 가상 서점을 그 대상으로 구축하였다.

본 논문의 구성은 2장에서 상품 정보 제공 방법을 살펴 보고, 3장에서는 본 논문에서 제시하는 상품 검색 지원 에이전트를 소개하며 4장에서 상품 검색 지원 에이전트의 개발 방법을 보이고 5장에서 결론을 내린다.

2. 상품 정보 제공 방법

전자상거래를 위한 쇼핑몰에서는 고객 즉 구매자에게 구매하고자 하는 상품을 제공할 수 있는 시스템이 필수적이다. 기존의 쇼핑몰에서는 상품정보를 제공하기 위해 통상 인터넷 검색 엔진이 제공하는 기능과 유사한 전자 카탈로그를 제공하고 있다. 그 첫째가 주제별 상품정보의 제공이다. 이는 상품정보를 계층적으로 분류하여 해당 제품에 대한 정보를 하향식으로 얻을 수 있게 한다. 둘째는 키워드 또는 검색어에 의한 상품 정보의 제공이다. 구매자가 구매하려는 상품에 대한 검색어 제공하면 그 검색어와 관련한 상품 정보를 제공한다. 마지막으로 구매자의 구매 의도와 관련한 기타 정보의 제공이다. 이는 그 상품을 구입한 과거 구매자의 통계치를 제공하는 등 다양한 방법이 있다[4].

또한 인터넷 쇼핑몰이 전통적인 채널과의 경쟁에서 우위를 확보하기 위해서는 이용고객의 가치를 창조해 나가야 한다[5]. 즉, 구매자에게 다양한 부가가치를 제공해야 한다. 인터넷 쇼핑몰에서 제공할 수 있는 부가가치로는 여러가지가 있다. 그 중 실세계의 매장의 영업 사원과 마찬가지로 구매자에게 원하는 상품 외에 구매할 만한 새로운 상품을 소개하는 것은 판매를 유도하며 구매자에게 정보를 제공하는 중요한 서비스이다[4].

본 논문에서의 대상 영역인 인터넷 가상 서점의 경우, 국내 사이트 예를 들면 교보 문고(http://www.kyobobook.co.kr/), 종로 서적(http://club.shopping.co.kr/seoul/book/), 와우북 (http://wowbook.com/), 인터파크(http://www.interpark.com/) 등에서는 주제별 검색이나 단순한 검색어를 이용한 검색만을 지원하고 있다 즉, 단순히 인터넷을 통한 서적의 판매에만 치중하고 있다. 국외 사이트의 경우 인터넷 가상 서점으로 가장 잘 구현되고 성공한 사례로 아마존(http://www.amazone.com/)을 들 수 있다. 그러나 아마존의 경우도 관련 상품에 대한 검색어로 단순히 상품에 대한 세부 사항 입력시 행해졌던 관련 상품의 분야와 해당 상품의 판매 이력만을 제공하고 있다[6]. 이러한 방식은 시시각각 변화하고 있는 구매자의 검색 지원요구를 제대로 반영하지 못하며 구매자가 필요한 상품을 제시하는데 부족함이 있다.

예를 들어 아마존에서 XML을 검색어로 검색할 경우 아마존의 자체 검색 엔진은 다음과 같은 순서로 검색 결과를 보여준다.

1) 검색어에 해당하는 베스트 셀러 목록
2) 검색어에 해당하는 서적 목록

여기서 제시되어진 서적을 선택할 경우 아마존 검색 엔진은 해당 서적의 상세 설명과 함께 단순히 그 서적이 속한 분야를 제시함으로써 검색어를 제안한다. 예로써 XML 관련된 베스트 셀러 목록 중 하나를 선택하면 다음과 같은 검색어를 제안한다.

1. XML (Document markup language)
2. Computer Programming Language
3. Computer Networks
4. Computer Bks - Communications / Networking
5. Computers
6. Programming Languages - General
7. Internet - Web Site Design

그러나 이것은 XML과 관련있는 분야라기 보다는 단순히 검색어의 결과로 나온 서적의 분야를 계층적으로 제시하는 것에 불과하다. 실제로 XML에 관심있는 구매자라면 HTML, 전자 상거래, 심지어 Java와 같은 분야에도 관심이 있을 것이다. 즉, 아마존

인터넷 가상서점은 계층 구조에서 하위 개념들간의 관련성을 고려하고 있지 못하다.

또한 시대의 변화에 따라 관련 있는 분야 또한 변하기 마련이다. 즉, 상품 입력 시행해졌던 분야에 근거한 관련 분야 제시는 구매자의 욕구를 제대로 충족시키지 못할 뿐 아니라 변화에 적응하기가 쉽지 않다.

본 논문에서는 이와 같이 제품의 해당 분야에 의존하여 상품 정보로써 관련 상품에 대한 관련 분야를 제시하는 대신, 구매자의 사이트 탐색 행위를 분석함으로써 적절하면서도 적시에 각 상품에 대한 관련 분야를 제시하는 상품 검색 지원 에이전트를 제시하고 개발한다.

3. 상품 검색 지원 에이전트

에이전트란 사용자를 대신해서 사용자가 원하는 작업을 자동적으로 해결하여 주는 소프트웨어이다. 따라서 본 논문에서는 상품 검색 지원 에이전트를 구매자를 대신해서 구매자가 원하는 것과 관련한 상품에 관한 정보를 자동적으로 제공해 주는 소프트웨어로 간주한다. 이러한 에이전트의 중심이 되는 개념은 자율성(autonomy), 적응성(adaptation), 협동성(cooperation) 등이다[7].

자율성은 사용자나 다른 에이전트의 직접적인 지시 없이도 스스로 행동하는 성질이며 에이전트를 특징 지워주는 가장 두드러진 성질이다[4]. 자율성에는 환경의 변화에 수동적으로 대처하는 반응성(reactivity) 외에도 자신의 목적을 가지고 능동적으로 작업 수행을 추구하는 목적 지향성(goal-directed behavior)도 포함된다. 적응성은 학습할 수 있는 능력을 나타내며 환경이나 사용자에 맞도록 적응하여 수행 효율을 높이는 성질이다.

에이전트는 그 특성상 정보 검색이나 정보 여과등의 인터넷 정보 가공을 위해 쉽게 적용될 수 있는 기술이다. 본 논문에서 구현되어진 상품 검색 지원 에이전트는 자율성과 적응성을 가지며 상품에 관한 정보를 사용자의 구미에 맞도록 가공하고 걸러주는 정보 여과 에이전트이다. 본 장에서는 상품 검색 지원 에이전트가 작동하는 개념과 기본 원리에 대해 살펴 본다.

구매자가 일반적인 인터넷 가상 서점을 방문했을 때 발생할 수 있는 시나리오는 다음과 같다.

1. 원하는 서적의 목록을 검색어를 통해서 검색한다.
2. 검색 결과 중 관심있는 서적의 제목을 보고 상세 내역을 선택한 후 확인한다.
3. 위 1,2의 과정을 반복하여 원하는 서적을 찾는다.

그러나 본 논문에서 구현한 인터넷 가상 서점을 방문했을 때 발생하는 일반적인 시나리오는 다음과 같다.

1. 원하는 서적의 목록을 검색어를 통해서 검색한다.
2. 검색 결과 중 관심있는 서적의 제목을 보고 상세 내역을 선택한 후 확인한다.
3. 본 논문에서 구현한 상품 검색 지원 에이전트에서 제공하는 상세 내역을 확인한 서적과 관련 있는 주제어에 의한 새로운 검색을 시도한다.
4. 1,2,3의 과정을 반복하여 원하는 서적을 찾는다.

이와 같은 시나리오를 토대로 사용자의 탐색 행위에 따라 수집할 수 있는 정보는 아래와 같다.

첫째, 구매자가 한 번의 세션 동안 선택하여 상세 내역을 확인 한 서적의 순서.
둘째, 그 서적이 속한 분야.
셋째, 제시된 분야에 대한 구매자의 반응이다.

이러한 정보를 수집하고 분석하기 위해서 본 논문에서 구현한 상품 검색 지원 에이전트는 인터넷 가상 서점을 방문하는 구매자에 대해 다음과 같이 반응한다.

1. 구매자의 탐색 행위 추적
2. 구매자의 탐색 행위를 분석하여 개개의 서적과 관련된 분야들에 대한 연관도를 계산
3. 구매자에게 연관도에 의한 관련 상품 정보 제시

3.1. 각 서적과 관련 분야들간의 연관도

서적을 입력할 때 만들어진 각 서적의 분야는 입력할 때 사용되어진 정보를 이용하여 분야의 계층 구조를 가진다. 본 논문에서 제안하는 검색 지원 에이전트는 이러한 계층 구조와는 별도의 관련 분야 리스트를 가진다. 즉, 각 서적은 자신의 관련 분야 리스트를 가지고 있다. 그리고 리스트 상의 각 관련 분야는 해당 분야와의 연관도를 가진다. 이러한 연관도는 사용자가 그 분야를 선택할 때 마다 증가되는 연관도 변화량의 합으로 정의한다. 또한 이 연관도가

클수록 두 분야는 연관성이 크다고 본다. 결과적으로 구매자에게 제공되는 정보는 바로 이 구매자가 선택한 서적의 관련 분야 연관도별 순위가 된다. 아래 그림 1은 관련 분야 리스트, 연관도, 계층 구조의 관계를 도식화하고 있다.

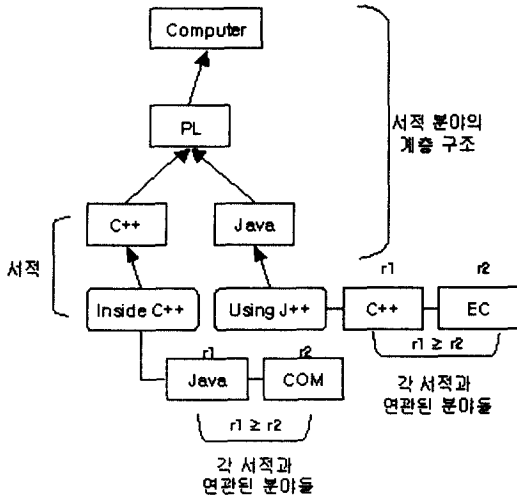



그림 1. 관련 분야 리스트, 연관도, 계층구조의 관계

이러한 연관도 변화량은 구매자의 서적 탐색 순서, 해당 분야의 계층 구조상 공통 분야, 포괄성 및 상품 검색 지원 에이전트가 제안한 관련 분야에 대한 구매자의 반응을 고려해서 조정되어진다.

3.2. 연관도 변화량

구매자가 탐색 행위도중 여러 종류의 서적을 선택했을 때, 각 서적들은 각기 자신의 분야를 가지고 있다. 그런데 구매자가 선택한 서적이 속한 분야들 사이의 연관도 변화량은 상호 비대칭적이다. 즉, XML을 선택한 사용자가 전자상거래에 대해 가지는 관심도와 XML을 선택한 사용자가 전자상거래에 대해 가지는 관심도는 서로 다르다.

이러한 비대칭적인 연관도 변화량을 고려하기 위해서 구매자의 탐색 순서, 각 서적이 포함되어져 있는 분야의 계층 구조, 각 서적이 포함되어져 있는 분야의 포괄성을 고려해야 한다.

3.2.1 구매자의 탐색 순서

일반적으로 인터넷 가상 서점에서 구매자의 탐색은 구매자가 가장 관심 있는 분야부터 시작한다. 구매자가 처음 입력한 검색어의 결과로 검색되어진 서적의 분야는 그 구매자가 해당 시점에서 가장 관심 있는 분야가 될 것이며 계속되어지는 탐색은 가장 관심 있는 분야와 관련 있는 분야를 찾기 위한 경우가 일반적이다. 또한 계속되어지는 탐색에서 만들어진 분야들의 리스트 각 원소 사이에도 일정한 정도의 연관성이 있다.

이와 같이 구매자의 탐색 순서에 따라 각 분야의 연관도를 판단하려면 구매자의 탐색 행위 도중 선택되어진 서적 분야의 순서 리스트에서 연관성을 가지는 각 분야의 위치를 찾아내는 것이 필요하다.

구매자가 인터넷 가상 서점을 이용할 때 선택한 서적의 분야를 발생하는 탐색 순서에 따라 아래와 같이 나누어 볼 수 있다.

1. 구매자가 최초 선택한 서적의 분야(*FC:First searched Category*)
2. 구매자가 계속해서 선택한 서적의 분야(*CC: Continuing searched Category*)
3. 제시된 분야 중 구매자가 검색하기 위해 선택한 분야(*SC:Successfully recommended Category*)

구매자의 탐색 순서에 따라 이와 같이 서적의 분야를 분류하는 이유는 첫째, 대부분의 쇼핑몰 사용자의 경우 사용자의 검색 순서는 찾고자 하는 서적부터 시작하여 자신이 원하는 서적을 발견하기까지 관련 있는 서적을 중심으로 검색을 하기 때문이며 둘째, 비록 구매자가 원하는 서적을 발견하여 계속되어진 검색이 실제로 전혀 다른 관점에서 검색되어진 것이라 할지라도 동일한 구매자가 가지는 여러 관심사들은 실제로 관계가 있기 때문이며 이러한 관련 있는 관심사를 추적하여 새로운 구매자들에게 정보를 제공하기 위함이다.

구매자는 *FC*, *CC*, *SC*와 관련된 탐색 행위를 반복하게 되는 데 이런 구매자의 탐색 행위를 단순화하기 위해서는 일정한 패턴이 필요하다. 구매자의 탐색 순서에 따라 발생하는 분야를 탐색 순서로 패턴화하면 아래와 같이 나타날 수 있다.

패턴 1. $FC \rightarrow CC^1 \rightarrow CC^2 \rightarrow \cdots \rightarrow CC^{n-1} \rightarrow CC^n$

패턴 2. $FC \rightarrow SC^1,\ CC^1 \rightarrow SC^2,\ CC^2 \rightarrow SC^3, \ldots,$
$CC^n \rightarrow SC^{n-1}$

FC : 구매자가 첫 번째 선택한 서적의 분야
CC^n : 구매자가 선택한 분야 중 첫 번째를 제외한 분야
SC^n : 본 시스템에서 제시한 분야 중 구매자가 선택한 분야
n : 구매자가 선택한 서적 분야의 순서

구매자는 일련의 책을 가상 서점에서 선택하게 되는 데 예를 들어 구매자가 The Java Programming Language, Introduction to CORBA를 선택 한 후, 본 시스템에서 제시한 COM이란 분야를 선택했으며 그 분야에 속하는 Inside COM을 선택하고 다시 The Java Network Programming을 선택했다면, 상품 검색 지원 에이전트는 구매자의 탐색 행위를 아래와 같은 두 개의 리스트로 나누어 저장한다.

리스트 1. (The Java Programming Language, Java) → (Introduction to CORBA, CORBA) → (Inside COM, COM) → (The Java Network Programming, Java)
리스트 2. (CORBA, Inside COM, COM)

본 논문에서는 위의 탐색 과정을 패턴화 하기 위하여 리스트 1에서의 Java와 같이 중복되는 패턴은 첫 번째 발생 시점을 제외하고 제거한다. 그러므로 위의 두 리스트에 저장되어진 탐색 과정은 다음과 같이 패턴화 되어 진다.

패턴 1. Java → CORBA → COM
패턴 2. CORBA → COM

그러므로 리스트 1은 패턴 1에 의해 다음과 같이 변경되어진다.

리스트 1. (The Java Programming Language, Java) → (Introduction to CORBA, CORBA) → (Inside COM, COM)

그리고 위의 구매자 탐색 과정에서 고려되어되어져야 할 관계의 종류는 위에서 제시한 바와 같이 패턴 1에 의해 (Java, CORBA), (Java, COM), (CORBA, Java), (COM, Java), (CORBA, COM)이 되며, 패턴 2에 의해 (CORBA, COM) (COM, CORBA)가 된다.

여기서 패턴 1의 경우 일련의 탐색 행위에서 발생한 분야간의 관계의 종류는 (FC,CC^1), (FC,CC^2), ...,(FC,CC^n), (CC^n,FC), (CC^{n-1},FC),(CC^{n-2},FC),..., (CC^1,FC),...,(CC^1,CC^n),(CC^n,CC^1)와 같이 ${}_nP_2$의 순서쌍이 된다. 이러한 패턴 1은 이후 각 분야 간의 연관도 변화량을 결정하는 데 사용되어 진다.

패턴 2의 경우 고려해야 될 분야간 관계의 종류는 (FC,SC^1), (CC^1,SC^2), (CC^2,SC^3) ,..., (CC^n,SC^{n-1}), (CC^{n-1},SC^n) ,..., (SC^3,CC^2), (SC^2,CC^1), (SC^1,FC) 이다. 이 패턴은 제시한 결과에 대한 검증에 사용되어 질 수 있다. 이때 패턴 2를 구성하는 중요한 요소는 구매자가 본 시스템이 제시한 분야를 선택했을 때의 분야와 그 분야이전에 구매자가 마지막으로 탐색한 서적의 분야가 된다. 이러한 패턴 2는 패턴 1에서 결정된 연관도를 검증하는데 사용된다.

패턴 1에서 각 분야간의 연관도 변화량을 고려하기 위해서 본 논문에서는 다음과 같이 각 분야를 구매자의 탐색 순서 상의 위치와 각 분야간의 탐색 순서에서의 거리로 것을 제안한다.

첫째, FC로 시작하는 원소의 집합

$$R0 = \{FC \rightarrow CC^1,\ FC \rightarrow CC^2,\ \ldots,\ FC \rightarrow CC^n\}$$

둘째, FC로 종료하는 원소의 집합

$$R1 = \{CC^1 \rightarrow FC,\ CC^2 \rightarrow FC,\ \ldots,\ CC^n \rightarrow FC\}$$

셋째, 첫째와 둘째를 제외한 노드간의 길이가 1인 원소들의 집합

$$R2 = \{CC^1 \rightarrow CC^2,\ CC^2 \rightarrow CC^1,\ \ldots,\ CC^{n-1} \rightarrow CC^n\}$$

....

N번째, 첫째와 둘째를 제외한 노드간의 길이가 N-2인 원소들의 집합

$$RN = \{CC^1 \rightarrow CC^{n-1}\}$$

여기서 각 집합의 원소들의 연관도 변화량은 위에서 제시된 집합에 따라 적용되는 기준을 달리한다. 따라서 위에서 제시되어진 집합의 순서에 따라 적용되는 연관도 변화량은 $RF0_i$, $RF1_i$, $RF2_i$, $RF3_i$이 된다. 즉, $RF0_i$와 $RF1_i$를 제외한 연관도 변화량은 구매자가 선택한 서적의 분야들 간의 구매자가 선택한 분야의 리스트에서의 거리에 따라 달라진다. 여기서 $RF0_i$와 $RF1_i$을 따로 분리하는 이유는 위에서 언급한 바와 같이 구매자가 처음 쇼핑몰에서 검색한 서적

의 분야와 바로 그 다음 검색한 서적의 분야사이의 연관도 변화량을 다르게 고려하기 위함이다.

예를 들면, FC로 시작하는 원소의 집합($R0$)에서 발견되는 관계에 대한 변화량의 기준은 $RF0_i$가 되므로 집합 내의 각각의 관계에 대한 변화량의 순서에 따라 $RF0_1, RF0_2, RF0_3, RF0_4, \ldots, RF0_n$이 될 것이다.

위의 RFN_k는 패턴 1내에서의 두 분야의 거리, 각 분야가 패턴에 나타난 순서에 의해 결정되어 진다. 즉, 관계를 구하고자 하는 두 분야가 C_i, C_j라고 할 때 본 논문에서 제시한 RFN_k는 다음과 같은 식으로 계산되어 진다.

$$RFN_k = \frac{1}{d_{i_j}} \quad (1)$$

여기서, $k \geq 2$이다. 그러나 k = 0, 혹은 k = 1 인 경우는 구매자가 가상 서점을 방문했을 때 처음 선택한 서적이므로 구매자가 가상 서점을 방문한 처음 의도를 반영할 수 있도록 $\frac{1}{d_{i_j}}$ 대신 적당한 상수로 대신한다.

3.2.2 해당 분야의 포괄성

구매자가 탐색 과정의 결과로 선택한 분야들은 서로 다른 범위의 포괄성을 가질 수가 있다. 이러한 포괄성에 따라 연관도 변화량이 달라질 수 있다. 즉, XML과 전자상거래의 연관도 변화량은 그 분야의 포괄성에 따라 구매자의 탐색 순서와는 다른 관점에서 산출되어 진다. 그러나 이러한 포괄성은 아마존에서 제공하고 있는 것과 같은 분야의 계층 구조만으로는 판단할 수 없다.

3.2.2.1 계층 구조상 완전하게 상,하위 개념일 경우

만약 구매자가 탐색 과정의 결과로 선택한 서적의 분야가 Java, 프로그래밍 언어라면 이 두 분야는 아래 그림 2와 같은 계층 구조를 가질 것이다.

이 경우 하위 분야는 상위 분야에 대해 모든 변화량을 무시한다. 그러한 경우 두 분야의 연관성은 이미 계층 구조에 잘 나타나있기 때문이다. 그러나 상위 분야로부터 하위 분야로의 연관도 변화량은 별도의 고려를 하지 않는다. 그 이유는 구매자의 탐색 행위를 분석함으로써 현재 상위 분야에서 속하는 많은 하위 분야들 중 당시 구매자들이 관심 있어 하는 하위 분야를 유추할 수 있기 때문이다. 예를 들어 일반적으로 프로그래밍 언어 분야의 서적을 검색하는 구매자들이 Java 분야의 서적을 찾는 일이 많다면 현재 프로그래밍 언어 분야에서 Java를 가장 많이 선호함을 유추할 수 있다.

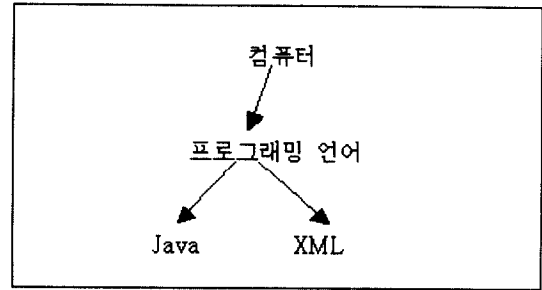



그림 2. Java, 프로그래밍 언어인 경우 두 분야의 계층 구조

3.2.2.2 계층 구조상 완전하게 상,하위 개념이 아닌 경우

구매자가 찾고자 하는 분야들은 그림 2의 XML, Java같이 대부분 3.2.2.1절에서 제시한 것처럼 계층 구조상 완전하게 하위 개념이기 보다는 대부분 계층 구조상에서 같은 수준의 개념이 될 것이다. 이러한 경우 포괄성은 단순히 분야들 간의 계층 구조로 판단하기는 매우 어렵다.

이러한 포괄성을 판단하기 위해서 계층 구조상 완전하게 하위 개념과 상위 개념으로 나누어지지 않는 분야들은 각 분야의 이미 만들어진 연관성을 이용한다. 즉, 연관된 분야가 더 많은 분야는 그렇지 않는 분야보다 더 포괄적이라고 본다.

만약 계층 구조가 완전하게 상,하위 개념이 아닌 두 분야 C^i와 C^j가 있다면 각 분야의 관련된 분야 개수와 그 분야들의 연관도를 고려하여 서로 얼마나 더 포괄적인지를 나타내는 포괄도비를 구할 수 있다. 만약 두 분야의 포괄도비를 계산하는 함수가 $g(C^i, C^j)$라면 FC와 CC^i의 포괄도비는 $g(FC, CC^i)$가 되며, CC^i와 FC의 포괄도비는 $g(CC^i, FC)$라고 한다. 즉, 이러한 포괄도비를 구해서 3.2.1절에서 구한 식 (1)에 적용하면 (FC , CC^i)의 변화량은 $RF1_i \cdot g(FC, CC^i)$가 된다.

여기서 관계를 구하고자 하는 두 분야의 포괄성을 의미하는 $g(C^i, C^j)$의 경우는 다음과 같은 식으로 계산되어 진다.

$$g(C_i, C_j) = \frac{RC(C_i)}{RC(C_i) + RC(C_j)} \qquad (2)$$

여기서, $RC(C_j)$는 본 시스템이 설정한 관계 중 C_j분야와 연관되어진 분야의 수이다. 일반적으로 $RC(C_i) \neq RC(C_j)$ 이다.

3.2.3 공통적인 분야에 의한 변화량

그림 2와 같이 한 서적의 분야는 다른 서적의 분야와 계층구조상 많은 공통 부분을 가질 수 있다. 이러한 공통적인 카테고리의 교집합 원소의 개수에 따라 새로운 기준의 변화량을 부여할 수 있다. 즉, 고려하고자 하는 두 분야와 관련된 분야의 전체 개수에 대한 두 분야 의 공통 분야의 비를 $I(Ci,Cj)$ 라고 하고 식 (1)과 식 (2)에 적용하면 (FC , CC^i)의 연관도 변화량은 $RF1_i \cdot g(FC, CC^i) \cdot I(FC, CC^i)$가 된다.

여기서 두 분야의 공통 분야 비를 가리키는 $I(C_i, C_j)$의 경우는 다음과 같은 식으로 나타낼 수 있다.

$$I(C_i, C_j) = \frac{H(C_i, C_j)}{H(C_i) + H(C_j)} \qquad (3)$$

여기서, $H(C_i)$는 C_i분야의 상위 분야의 개수이며, $H(C_i, C_j)$는 C_i와 C_j의 공통 분야 중 두 분야에 공통적인 분야의 개수이다.

3.3. 연관도 변화량 계산

위의 식 (1), 식 (2), 식 (3)을 바탕으로 연관도 변화량을 계산 할 경우 연관도 변화량은 아래와 같다.

연관도 변화량 (Δr) =

$$\omega \cdot RFN_k \cdot g(C^i, C^j) \cdot I(C^i, C^j) \qquad (4)$$

(단, $0 < RFN_k \cdot g(C^i,C^j) \cdot I(C^i,C^j) < 1$)

그리고 구매자가 3.2.1절에서 제시한 패턴 2의 탐색 행위를 할 경우 이것을 설정된 관계의 검증에 사용한다.

본 시스템에서 제시한 분야를 사용자가 검색하기 위한 검색어로 사용했다면 역시 그 분야의 연관도를 상향 조정한다. 이 경우 연관도 변화량 $RFN_k \cdot g(C^i,C^j) \cdot I(C^i,C^j)$ 은 1이 된다. 또한 패턴 2에 속하지 못한 분야일 경우 두 분야의 연관도 변화량은 $-1 < RFN_k \cdot g(C^i,C^j) \cdot I(C^i,C^j) < 0$인 임의의 상수로 정의한다. 이 것은 두 관계의 연관도가 계속해서 증가하는 것을 방지하기 위해서이다.

또한 먼저 사용자가 서적을 선택한 후, 그 관련 분야를 선택할 확률이 한계치(PT) 이하이고 그 분야의 연관도 역시 특정 한계치(RT)보다 작으면 그 분야를 제거한다. 이러한 한계치들은 검색 지원 에이전트의 시스템 부하와 구매자의 편의성을 고려하여 설정한다.

3.4. 연관도 계산

3.3절의 연관도 변화량 식 (4)를 토대로 계산되어지는 연관도(R)은 다음과 같다.

$$R = \Sigma(\Delta r) (\text{단 } 0 < R < 1) \qquad (5)$$

$$(\text{단, } \Delta r = \omega \cdot RFN_k \cdot g(C^i, C^j) \cdot I(C^i, C^j))$$

즉, 전체 연관도는 구매자의 탐색 행위 분석 시 도출되어지는 연관도 변화량(Δr)의 총 합이 된다.

3.5. 상품 검색 지원 에이전트의 동작 예

구매자는 구현한 인터넷 가상 서점 사이트에 머물면서 한 세션 동안 비정규적인 탐색 활동을 한다. 이를 분석하여 3.2.1절에서 제시한 두 가지 패턴으로 단순화하는 것이 필요하다. 그러므로 사용자가 검색한 탐색 행위를 기록한 후, 아래와 같은 별도의 단순화 과정을 거쳐야 한다.

첫째, 한 구매자가 탐색한 분야의 리스트에서 반복적으로 나타나는 분야는 해당 분야의 최초 발생 시점만을 남기고 생략한다. 이 때 반복 정도에 따라 연관도 변화량을 조정할 수 있다.

둘째, 패턴 1과 패턴 2를 필터링하여 이를 분리하여 처리한다.

셋째, 구매자는 구현한 인터넷 가상 서점에서 동일 세션동안 하나의 관심 분야에 관련된 것만을 탐색하는 것은 아니다. 즉, XML을 검색한 후, 전혀 다른 관점에서 VRML을 검색할 수 있다. 또한 구매자의 일련의 행위를 분석하는 계산의 과부하를 막기 위해서 모든 탐색 과정을 조사하지 않고 일정 길이이하만을 조사한다.

만약 구매자가 해당 분야가 CORBA인 서적을 선택했다면 본 검색 지원 에이전트는 사용자에게

CORBA 관련 분야 목록을 연관도 순위별로 제시한다. 구매자는 계속해서 제시된 목록 중 관련 분야가 Java인 것을 선택한 후 그 중 관련 분야가 COM인 책을 선택했다. 다시 사용자는 탐색 도중 해당 분야가 CORBA인 서적을 선택했다. 이러한 시나리오에서 본 검색 지원 에이전트는 3.2.1절에서 제시한 패턴 1을 이용하여 CORBA → COM의 리스트를, 패턴 2를 이용하여 CORBA → Java의 리스트를 만든다. 이 경우 CORBA → COM의 연관도 변화량은 $RF0_1 \cdot g(\mathrm{CORBA,COM}) \cdot I(\mathrm{CORBA,COM})$이 되며 COM → CORBA의 연관도는 $RF0_1 \cdot g(\mathrm{COM, CORBA}) \cdot I(\mathrm{COM,CORBA})$이 된다. 이때 CORBA와 연관된 분야가 Java, C++, OOP, 네트워크, UNIX라고 하고 COM과 관련된 분야가 VB, C++, Java, Windows라고 하면 $g(\mathrm{CORBA,COM})$ 는 5/9, $g(\mathrm{COM, CORBA})$는 4/9가 된다. 또한 COM이 속한 분야의 계층 구조가 Computer, Distributed Object, DCOM이고 CORBA가 속한 분야의 계층 구조가 Computer, Distributed Object라고 하면 $I(\mathrm{COM, CORBA})$는 2/6이 된다. 이러한 연관도 변화량을 각각 CORBA의 COM 연관도, COM의 CORBA 연관도에 반영한다. 또한 패턴 2를 이용하여 CORBA의 Java 연관도 역시 증가 시키며 CORBA의 관련 분야 중 3.3절에서 제시한 한계치(PT, RT)를 넘지 못하는 분야는 관련 분야 리스트에서 제거한다.

4. 상품 검색 지원 에이전트의 개발

본 논문에서 구현한 상품 검색 지원 에이전트의 개발 환경은 운영체제로 윈도우NT 4.0, 웹 서버로 IIS 4.0이며 개발 도구는 Visual Basic 6.0 WebClass, VC++ 6.0 ATL 2.0, Visual InterDev 6.0을 사용하였으며 데이터베이스로는 SQL 서버7.0을, 필요한 서적 데이터를 모으기 위한 서치 엔진 제작을 위하여 WinInet API를 사용하였다. 본 논문에서 제작한 에이전트는 특히 전자 상거래 서버와 웹 서버의 부하를 분산하기 위하여 DCOM을 사용한 독립적인 어플리케이션 서버로 구현되었다. DCOM의 장점은 언어와 플랫폼에 독립적인 컴포넌트 단위의 객체라는 것 외에도 부하 분산을 들 수 있는데, 현재로써는 위의 모든 서버는 한 시스템에 설치되어 있지만, 간단한 설정 변경을 통하여 전용 상품 검색 지원 에이전트 서버로 부하를 분산할 수 있다[8].

다음 그림 3은 제안한 인터넷 쇼핑몰을 위한 상품 검색 지원 에이전트를 포함하는 인터넷 가상 서점의 전체 시스템 구성도이다.

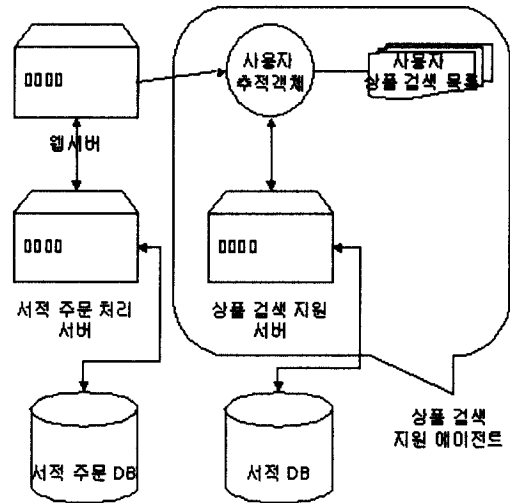



그림 3. 상품 검색 지원 에이전트를 사용한 인터넷 가상 서점 구성도

그림 3의 사용자 추적 객체는 사용자 별로 관리되며, 사용자가 웹 서버에 처음 접근하여 해당 웹을 떠나는 동안 사용자의 상품 검색 목록을 기록한다. 사용자의 입력이 일정한 시간동안 발생하지 않으면 즉, 한 세션이 종료하면, 사용자 추적 객체는 소멸되며, 관리하고 있던 사용자 검색 목록은 상품 검색 지원 에이전트로 전달된다. 상품 검색 지원 에이전트는 사용자의 입력을 패턴화하여 본 논문에서 제시한 연관도 변화량을 계산하여 각 분야들의 연관도를 산출하게 된다. 또한 구매자가 3.2.1절에서 제시한 패턴 2의 탐색 행위를 할 경우 이것을 설정된 관계의 검증에 사용한다.

4.1. 상품 검색 지원 에이전트가 사용하는 데이터 베이스 스키마

본 논문에서 구현한 상품 검색 지원 에이전트는 각 서적에 관한 연관도 리스트를 저장하기 위하여 기존의 서적 DB에 하나의 새로운 테이블과 몇 개의 추가적인 속성을 추가해 사용한다. 서적DB에서 본 논문에서 구현한 검색 지원 에이전트와 관계 있는 데이터 베이스 스키마는 다음과 같다.

Category (**category_id**, category_name, parent_cat,...)
Books (**book_id**, category_id, title, writer, isbn,...)
RelationShip (**book_id, category_id**, relation_rank,...)
...

이 스키마에서에서 Category의 parent_cat는 분야의 계층 구조를 나타내며, RelationShip의 book_id와 category_id는 서적과 그 관련 분야를 의미하며, relation_rank는 해당 book_id의 서적과 category_id가 참조하는 분야간의 연관도를 나타낸다. 즉, RelationShip과 Books간의 관계와 RelationShip과 Category간의 관계는 모두 다대일 관계가 된다.

4.2. 사용자 추적 객체의 구현

사용자 추적 개체의 목표는 구매자가 가상 서점에서 서적을 탐색하는 과정을 기록하고 구매자의 탐색 과정이 종료하면 즉, 가상 서점을 떠날 경우 상품 검색 지원 서버에게 구매자의 탐색 과정을 전달하고 구매자의 탐색 패턴 분석을 지시하는 것이다. 본 논문에서 구현한 사용자 추적 객체는 ASP환경의 구매자가 처음 가상 서점에 접속 시 발생되는 Session객체의 Start 이벤트 핸들러내에서 생성된 후 구매자의 한 세션동안 생명 주기를 가지는 Session 변수에 저장되어진다. 또한 구매자가 가상 서점을 떠날 경우 실행 될 Session객체의 End 이벤트 핸들러에서 구매자 탐색 과정을 상품 검색 지원 서버로 전달한 후, 사용자 추적 객체를 소멸한다. 이때 사용되어 지는 상품 검색 지원 서버에 대한 참조는 가상 서점 웹 응용 프로그램이 처음 실행될 때 발생되어지는 Application객체의 Start 이벤트 핸들러내에서 처음 생성되어지고 가상 서점 웹 응용 프로그램과 동일한 생명 주기를 가지는 Application객체에 저장되어 진다.

다음 그림 4는 상품 검색 지원 에이전트의 사용자 추적 객체 및 상품 검색 지원 서버에 대한 생성 및 초기화를 담당하는 ASP Script 코드이다.

사용자의 탐색 행위를 기록하고 추적하기 위해서는 구매자의 서적 선택 시 활성화될 이벤트가 필요하다. 본 논문에서는 WebClass의 WebItem이벤트를 이용한다. 즉, 구매자가 서적을 선택할 때마다 발생할 WebItem이벤트를 등록하고 이에 해당하는 이벤트 핸들러를 작성한 후 등록한 이벤트 핸들러내에서 구매자의 탐색 순서와 각 서적의 선택 시, 선택한 서적과 그 서적의 분야를 기록한다[9]. 이때 선택한 서적과 그 서적의 분야는 HTTP 프로토콜의 Request 메시지의 GET 메쏘드를 이용한다[10]. 또한 패턴 1과 패턴 2를 구분하기 위해서 사용자의 검색이 서적 검색 지원 에이전트가 제시한 분야에 의한 것인지를 구별하여 기록한다. 그리고 패턴 2를 작성하기 위하여 마지막으로 구매자가 선택한 서적의 분야를 Session 변수에 저장해야 한다. 다음 그림 5는 구매자의 서적 선택 시 실행되어질 이벤트 핸들러내에서 탐색의 순서와 서적, 서적의 분야를 기록하는 WebClass 프로그램이다.

```
' Session객체의 Start 이벤트 핸들러
Sub Session_Start()
' 사용자 추적 객체의 생성 후
' Session 변수에 저장
  Set Session("CustomerTrack") =
    CreateObject("CustomerSearchList")
End Sub
' Session객체의 End 이벤트 핸들러
Sub Session_End()
' 검색 지원 서버에 관한 객체 참조
 Application("aSrchAgentSrv").Pattern _
   (Session("CustomerTrack"))
End Sub
' Application객체의 Start 이벤트 핸들러
Sub Application_Start()
  ' 상품 검색 지원 서버에 대한 참조 생성
  Set Application("aSrchAgentSrv") =
     CreateObject("aSrchAgentSrv")
End Sub
```

그림 4. 사용자 추적 객체 및 상품 검색 지원 서버의 생성 및 초기화

4.3. 상품 검색 지원 서버의 구현

다음 그림 6은 상품 검색 지원 에이전트 서버의 내부 구성이다. 구매자 탐색 과정 추적 객체는 구매자의 탐색 행위를 해당 구매자의 세션동안 관련 분야 제시에 의한 검색과 검색어 입력에 의한 검색을 나누어서 기록한다. 구매자의 세션이 종료되는 순간, 추적 결과 패턴화 객체는 기록된 결과를 3.2.1절에서 제시한 구매자 탐색 순서 패턴으로 단순화된다. 만약 사용자의 탐색 행위가 구매자의 검색어에 의한 것이

```
' 구매자가 서적을 선택할 때
' 실행 될 이벤트 핸들러
Sub Category_Select
  If Request.QueryString("suggest") = "" Then
    ' 구매자가 선택한 서적이 본 시스템에서
    ' 제시한 분야에 의한 것이 아닐 경우
    Session("CustomerTrack").AddCategoryFC
    (Request.QueryString("book_id"),
        Request.QueryString("category_id"))
  ' 패턴 2를 위하여 현재 서적의 분야를 저장
    Session("prev_category_id") =
        Request.QueryString("category_id")
  Else
    ' 구매자가 선택한 서적이 본 시스템에서
    ' 제시한 분야에 의한 것일 경우
    Session("CustomerTrack").AddCategorySC
    (Request.QueryString("book_id"),
        Request.QueryString("category_id"),
    Session("prev_category_id"))
    ' 본 시스템에서 제시한 분야를
    ' 선택하기 전에 마지막으로 구매자가
    ' 선택한 서적의 분야를 함께 저장
    Session("prev_category_id") = ""
  End If
End Sub
```

그림 5. 서적 탐색의 순서, 각 탐색 시 서적, 서적의 분야를 기록하는 WebClass 프로그램

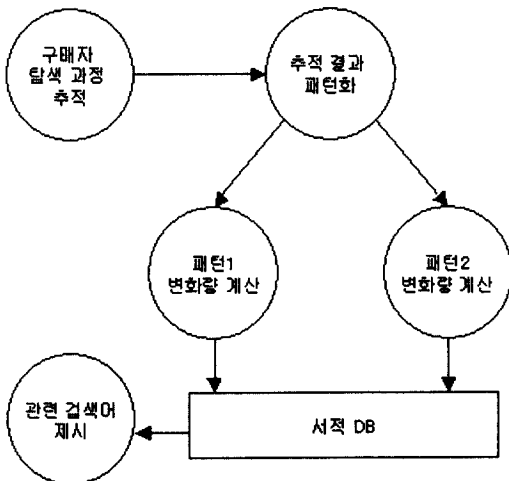



그림 6. 상품 검색 지원 에이전트 서버의 구성

라면 패턴 1 변화량 계산에 의해, 3.3절에서 제시한바와 같이 연관도 변화량을 계산하게 되고, 상품 검색 지원 에이전트에서 제시한 관련 분야에 의한 것이라면 패턴 2 변화량 계산에 의해 3.3절에서 제시한 설정된 관계 검증에 의해 계산되어진다. 이 결과는 저장되어진 후, 본 논문에서 구현한 가상 서점을 방문하는 다른 구매자에게 제시되어진다.

3.3절에서 제시되어진 연관도 변화량은 구매자가 탐색 행위를 마칠 때, 상품 검색 지원 서버에서 해당 구매자의 탐색 순서를 분석하여 패턴화 한 후 일괄적으로 적용하게 된다. 본 논문에서는 각 서적의 연관도 리스트를 유지하기 위하여 4.1절의 상품 검색 지원 에이전트의 데이터 베이스 스키마를 이용하였으며, Visual C++의 ATL과 ADO를 사용했다. 4.1절의 상품 검색 지원 에이전트의 데이터 베이스 스키마를 이용하여 패턴 1과 패턴 2에 의해 고려해야 될 모든 관계에 대해 적용될 상품 검색 지원 에이전트의 연관도 리스트 갱신 프로그램은 대략 다음 그림 7과 같다. 아래 프로그램은 MS SQL 7.0에서 저장 프로시저로 작성되어졌으며 ADO를 통해서 호출되어 진다.

```
IF EXISTS(SELECT * FROM RelationShip
 WHERE Book_id =
 (구매자가 선택한 서적의 book_id) AND
 category_id =
 (고려해야 할 관계 (Ci,Cj) 중
   Ci에 해당하는 category_id))
THEN
  UPDATE RelationShip
  SET RelationRank = RelationRank + Δr
  /* Δr = RFNk · g(Ci,Cj) · I(Ci,Cj) */
  FROM RelationShip INNER JOIN Books
  ON RelationShip.book_id = Books.book_id
  INNER JOIN Category
  ON RelationShip.category_id =
                  Category.category_id
  WHERE RelationShip.book_id =
  (구매자가 선택한 서적의 book_id) AND
  RelationShip.category_id =
  (고려해야 할 관계 (Ci,Cj) 중
    Ci에 해당하는 category_id)
ELSE
  INSERT INTO RelationShip
  (book_id, category_id, relation_rank)
  VALUES (
  (구매자가 선택한 서적의 book_id),
  (고려해야 할 관계 (Ci,Cj) 중
    Ci에 해당하는 category_id), Δr)
  /* Δr = RFNk · g(Ci,Cj) · I(Ci,Cj) */
END IF
```

그림 7. 연관도 리스트 갱신 알고리즘

다음 그림 8은 본 논문에서 구현한 상품 검색 지원 에이전트에서 구매자가 CORBA관련 서적을 선택하여 상세 내역을 확인할 때 제시되는 분야를 보여 주고 있다.

그림 8. 본 시스템에서 제안하는 CORBA 관련 분야

위의 그림 8에서 선택한 서적의 경우 서적 자료 입력 시 서적의 분야는 계층 구조 순으로 CORBA, Object Oriented Programming, Programming-General, Computer Books, Computers이었다. 그러나 구매자의 탐색 패턴을 분석한 결과 위의 서적에 대한 관련 분야는 Java, COM+/DCOM, IDL과 같은 분야가 추가되었다. 이러한 서적 자료 입력 시에 예상할 수 없었거나 새롭게 대두된 분야들이다. 이러한 분야들은 현재 구매자들의 탐색 관심 사항을 반영하고 있으며 이러한 분야를 구매자들에게 제시함으로써 구매자들에게 보다 효과적인 구매정보를 제공하고 있음을 알 수 있다.

5. 결 론

인터넷 사용인구의 폭발적인 증가에 따라 인터넷상의 전자상거래가 일반화되고 있다. 전자상거래에서 기존의 쇼핑몰들은 상품 정보에 대한 주제별 검색 또는 검색어에 의한 상품 정보를 제공하고 있는 것이 일반적이다. 또한 보다 효과적인 상품정보를 제공하기 위해 특정 상품을 검색하는 구매자에게 해당 상품의 판매 이력을 제시함으로써 보다 효과적인 상품 정보를 제공하기도 한다. 그러나 구매자에게 보다 효과적인 구매 정보를 제공하려면 구매자의 구매 의도를 파악하여 보다 효과적인 구매정보를 제시하는 방법의 추구가 필요하다.

본 논문에서는 인터넷 가상서점을 대상으로 하여 쇼핑몰에서 보다 효과적인 구매 정보를 제공하기 위하여 과거 다른 구매자의 탐색 패턴을 분석하여 이를 근거로 하여 구매자가 구매하려는 상품과 관련한 상품 정보를 제시해 주는 상품 검색 지원 에이전트를 설계하고 구현하였다. 따라서 본 검색 지원 에이전트를 사용하는 시스템은 구매자에게 보다 효과적인 구매 정보를 제공할 수 있어 구매자는 보다 효율적으로 상품을 검색할 수 있다.

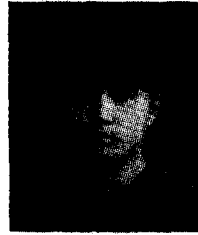
그러나 인터넷 쇼핑몰에서 전문화된 정보 제공 서비스를 제공하기 위해서는 많은 연구가 필요하다. 우선 본 논문에서 제안한 검색 지원 에이전트를 보다 효과적으로 구현하려면 기계학습을 통한 검색 지식베이스의 구축이 필요하다. 또한 상품 정보에 대한 보다 체계적인 상품 분류를 위한 분류 지원 시스템의 개발도 필요할 것으로 사료된다.

참 고 문 헌

[1] 주재훈, 인터넷 비즈니스 전자상거래, 비봉 출판사, 서울, 마포구, 1998.

[2] David Kosiur, Understanding Electronic Commerce, Microsoft Press, Radmond, Washington, 1997.

[3] G..Winfield Treese, Lawrence C. Stewart, Designing Systems for Internet Commerce, Addison Wesley, 1997.

[4] 최중민, "인터넷 정보 가공을 위한 에이전트", 정보처리 학회지, 제 4권 제 5호, pp.101-109.
[5] 최성, 이기민, "중소기업의 마켓팅 전략과 쇼핑몰 활용 연구," 정보처리학회지, 제6권 제 1호, pp.89-97.
[6] http://www.amazone.com/, Amazon.com Home.
[7] Alper Caglayan, Colin Harrison, "Agent Sourcebook", Wiley Computer Publishing, 1997.
[8] Guy Eddon, Henry Eddon, "Inside Distributed COM", Microsoft Press, Radmond, Washington, 1998.
[9] Sean Alexander, "The WebClass Developer's Primer", http://msdn.microsoft.com/library/techart/msdn_wbclsprmr.htm, 1998.
[10] R. Fielding, Hypertext Transfer Protocol-HTTP/1.1, RFC 2616, 1999.

홍 영 준

1998년 2월 동아대학교 컴퓨터 공학과 졸업 (공학사)
1999년 3월~동아대학교 컴퓨터 공학과 석사 과정
관심분야 : 전자 상거래, XML, 에이전트

박 경 환

1981년 경북대학교 전자과 전산 전공 학사.
1983년 서울대학교 컴퓨터공학과 석사.
1990년 서울대학교 컴퓨터공학과 박사.
1984~1986년 서울대학교 공과대학 시간강사.
1998년 University of California, Irvine 객원 교수.
1987~현재 동아대학교 전기전자컴퓨터컴퓨터공학부 교수.
관심분야 : 멀티미디어 시스템, 원격교육, 전자상거래, 가상현실.